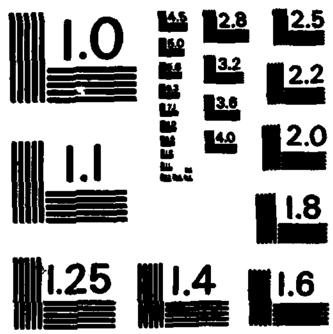
AD-A121 760 JACKKNIFE FOR VARIANCE ANALYSIS OF MULTIFACTOR EXPERIMENTS(U) NAVAL BIODYNAMICS LAB NEW ORLEANS LA R C CARTER ET AL. MAY 82 NBDL-82R013 1/1

UNCLASSIFIED F/G 12/1 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

NBDL - 82R013

Jackknife for Variance Analysis of Multifactor Experiments

Robert C. Carter
and
Alvah C. Bittner, Jr.

May 1982

NAVAL BIODYNAMICS LABORATORY
New Orleans, Louisiana

Approved for public release. Distribution unlimited.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER NBDL-82R013	2. GOVT ACCESSION NO. AD-A121760	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Jackknife for Variance Analysis of Multifactor Experiments		5. TYPE OF REPORT & PERIOD COVERED Research Report
		6. PERFORMING ORG. REPORT NUMBER NBDL-82R013
7. AUTHOR(s) Robert C. Carter and Alvah C. Bittner, Jr.		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Biodynamics Laboratory Box 29407 New Orleans, LA 70189		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS MF58524-02E-002
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Medical Research & Development Command Bethesda, MD 20014		12. REPORT DATE May 1982
		13. NUMBER OF PAGES 7
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Data Analysis, Statistical Jackknife, Variance

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A method is presented for analyzing effects of multi-factor experimental treatments on the variance of a dependent variable. The method is based on the statistical jackknife. It enables the analyst to enhance the power of an analysis by using the degrees of freedom associated with the random factor (e.g., subjects, in a behavioral experiment) of a multifactor design. The method is suitable for investigating hypotheses about trends of variances. A computer program is appended which calculates the jackknife variance estimates and other useful statistics.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SUMMARY PAGE

PROBLEM

Usually an experiment is conducted to show how the experimental treatments affect mean changes of some (dependent) variable. Sometimes there is interest in how the treatments affect the spread of the dependent variable observations.

FINDINGS

A method is presented for showing how the variability of the dependent variable is affected by treatments. The treatments may be defined in terms of several variables, effects of which can be isolated and evaluated. A FORTRAN language computer program is appended which implements the method.

RECOMMENDATION

Use the method described herein to show how multifactor treatments affect the variances of observations.

The work was funded by the Naval Medical Research and Development Command

Introduction

How was the variability of observations affected by experimental treatments? This question is often of interest during the analysis of an experiment. There are two typical reasons for such interest. First, intertreatment homogeniety of variances is an assumption required for analysis of effects of the treatments on the means of the observations (Winer, 1971). Therefore, a careful analyst assesses the credibility of the hypothesis of homogeniety of variance even though his primary interest is in changes of the means. In addition, sometimes the effect of the treatments on the variances of observations is the analysts' main interest (Bitter & Carter, 1982). This would be the case, for instance, if the treatment were repeated experience with some task (e.g., a mental test) and one wished to investigate the theory (Jensen, 1980) that the variance of observations increases with the amount of experience the subjects have had with the task.

Many methods are available for investigating the hypothesis of homogeniety of variances in experiments having only one independent variable (i.e., treatment) which is used at only a few levels of its range. These methods are summarized by Winer (1971) and Hollander and Wolfe (1975). Such methods are not directly applicable to experiments having multiple independent variables. In addition, they are not applicable for assessing trends in the variances associated with an independent variable whose range is sampled at many levels (e.g., Jensen's theory mentioned above implies an increasing trend in the variances as the independent variable of practice increases).

Games (1978) has suggested an approach to analysis of the variability of observations in multifactor experiments. In essence, he proposed that the natural log transformed (ln) variances be considered as observations in an experimental design having one observation per treatment condition. This ingenious idea, built upon earlier suggestions of Scheffe (1959), is adequate for many purposes. Its primary drawback is that it uses the interaction mean square to estimate its error variance. This use precludes investigations of the interaction. In addition, the strength of a test depends heavily on the number of degrees of freedom in the estimate of the statistical error (i.e., interaction mean square). If the experiment had $\underline{I}$ observations at each of $\underline{J}$ levels of one treatment by $\underline{K}$ levels of another treatment, Games' technique provides an error estimate with degrees of freedom equal to the product: $(\underline{K}-1)(\underline{J}-1)$. Hence the error estimate can be quite puisne if either $\underline{J}$ or $\underline{K}$ is small. Increasing the number of observations, $\underline{I}$, does not help directly because they are all collapsed into one observation (variance) per treatment condition. It appears that a better technique would somehow take greater advantage of the $\underline{I}$ observations per treatment condition. Then the statistical test of the variances would have adequate degrees of freedom if the number of observations were reasonable.

The statistical jackknife (Mosteller & Tukey, 1977) enables the analyst to use the degrees of freedom associated with the observations. It has been used to test homogeniety of variances (Hollander & Wolfe, 1973) in single-treatment experiments. Games, Keselman, and Clinch (1979) endorse its use in multifactor experiments.

The Jackknife

The statistical jackknife is, as the name implies, a handy general-purpose tool. The purpose for which it is used here is to approximate the distribution of variances in each treatment-condition of an experiment. According to Mosteller and Tukey (1977), the jackknife works well for this purpose as long as the data's sample distribution does not have straggling tails or abrupt end points, is roughly symmetrical, and has a sufficient number of observations.

The jackknife produces numbers which have a distribution like that of a combination of the original data (e.g., the variance is such a combination). If there are *I* observations available, then *I* variance surrogate numbers can be produced by the jackknife. These numbers can be analyzed to test the hypothesis of homogeniety of variance by using the same procedure used to test the hypothesis of homogeniety of the means of the original observations. The advantage of the unity of analysis made possible by the jackknife is difficult to overemphasize. Obviously, analysis of an experiment is simpler if the same procedure can be used (on the observations and the jackknife variances) to test for effects on the means and variances. Complete parallelism of the two analyses lends clarity of thought, in addition to simplicity of analysis. As an illustration of the type of analysis which can be done, the BMDP computer program P2V (Dixon, 1981) can be used to test multifactor experiments for main effects, trends, and interactions of means (and jackknife variances). To take advantage of this program for inferences on variances only jackknife variance estimates need to be prepared. A computer program will be presented which prepares jackknife variances for analysis.

A Computer Program to Prepare Jackknife Variances

The FORTRAN computer program of APPENDIX A produces the jackknife variance surrogate numbers referred to above. A jackknife variance (Hollande & Wolfe, 1977) is produced corresponding to each observation in each treatment condition (which may be defined by multiple treatment factors). Corresponding to observation number *i* in treatment *m* is

$$S_i^{*2} = N*\ln(S^2{}_{all}) - (N-1)*\ln(S^2{}_i)$$

where S^{*2} is a jackknife variance, $S^2{}_{all}$ and $S^2{}_i$ are the usual variance estimates calculated, respectively, with all observations in treatment *m* or all but observation *i* where *N* is the number of observations (e.g., subjects) in treatment *m*. The natural log-transformation (ln) is employed to reduce the effect of kurtosis on the dispersion of the variance estimates (Scheffe, 1959).

In APPENDIX A, $\underline{S}^2{}_{all}$ is taken from the diagonal of a variance-covariance matrix of all the observations in all the treatments, and $\underline{S}^2{}_i$ is a diagonal element of the variance-covariance matrix of all observations except the ith in each treatment condition. The variance-covariance matrix is generated by a subroutine named CORAN (UNIVAC, 1969). The jackknife variances are then punched on computer cards in the same format as the original data. The jackknife variances may then be analyzed conveniently using the same method used for the original data.

In addition, the program in APPENDIX A prints the original data and $\underline{S}^2_{all}$ for each treatment condition, and punches on computer cards the correlations among the observations in all pairs of treatment conditions. The original data are printed for identification and verification purposes. The variances, $\underline{S}^2_{all}$, for each treatment condition are printed for the information of the user. Furthermore, the correlations are punched for convenience of analysts who wish to explore the correlation structure of repeated measurements using a method such as that described by Steiger (1980). All of this provides the analyst with options not ordinarily available for more complete description of the data.

REFERENCES

Bittner, A. C., Jr., & Carter, R. C. Repeated measures of human performance: A bag of research tools. (Research Report No. NBDL-81R011) New Orleans, LA: Naval Biodynamics Laboratory, November, 1981. (NTIS No. AD-A113954)

Dixon, W. J. (Chief Editor) BMDP statistical software. Los Angeles: UCLA, 1981.

Games, P. A. A three-factor model encompassing many possible statistical tests on independent groups. Psychological Bulletin, 1978, 85, 168-182.

Games, P. A., Keselman, H. J., & Clinch, J. J. Tests for homogeneity of variance in factorial designs. Psychological Bulletin, 1979, 86, 978-984.

Hollander, M., & Wolfe, D. A. Nonparametric statistical methods. New York: Wiley, 1973.

Jensen, A. R. Bias in mental tests, New York: Free Press, 1980.

Mosteller, F., & Tukey, J. W. Data analysis and regression. Reading, Mass.: Addison-Wesley, 1977.

Steiger, J. H. Tests for comparing elements of a correlation matrix, Psychological Bulletin, 1980, 87, 245-251.

Scheffé, H. The analysis of variance. New York: Wiley, 1959.

UNIVAC, 1108 system stat pack, Minneapolis: Sperry Rand Co., 1969.

Winer, B. J. Statistical principles in experimental design. (2nd Ed.) New York: McGraw-Hill, 1971.

APPENDIX A

A FORTRAN Language Computer Program to Estimate Jackknife Variances

```
C      THIS IS A FORTRAN LANGUAGE COMPUTER PROGRAM TO CALCULATE JACKKNIFE
C      VARIANCE ESTIMATES AND OTHER USEFUL STATISTICS.  THE PROGRAM USES
C      CORAN, A UNIVAC STAT PACK SUBROUTINE TO CALCULATE CORRELATIONS AND
C      VARIANCES.  ANOTHER SUBROUTINE COULD BE SUBSTITUTED FOR CORAN.
       DIMENSION DATA(25,15),CORR(15,15),DCORR(15,15),DDATA(25,15),
      *PCOR(15,15),Y(375),SYMB(15),XBAR(15),DDDATA(3,25),(15,5)
      *PCOR(15,15),Y(375),SYMB(15),XBAR(15),DDDATA(3,25),15,5)
      *PCOR(15,15),Y(375),SYMB(15),XBAR(15),DDDATA(3,25,15,5)
C      NSUB,NDAY, ARE THE NUMBER OF LEVELS OF TWO INDEPENDENT VARIABLES.
       NSUB=15
       NDAY=15
       NK=1
       NT=NSUB
       NI=NSUB
       NJ=NDAY
C      ISUB,IDAY,IDEN ARE THREE INDEPENDENT VARIABLES.
       DO100ISUB=1,NSUB
       DO100IDAY=1,NDAY
       DO 100 IDEN=1,3
C      READ DATA
READ(5,200)(DDDATA(IDEN,ISUB,IDAY)
200    FORMAT (24X,F7.2,21XF7.2/24XF7.2,21X,F7.2/24X,F7.2)
100    CONTINUE
       DO 800 IDEN=1,3
C      WRITE DATA FOR IDENTIFICATION AND VERIFICATION.
       WRITE(6,91)
91     FORMAT(1H ,'DATA:')
       DO 400 KSUB=1,NSUB
       WRITE(6,25)(DATA(KSUB,IDAY),IDAY=1,15)
25     FORMAT(1H ,15F7.2)
400    CONTINUE
C      WRITE CORRELATION MATRIX WITH LEVELS OF IDAY AS THE INDEX OF THE
C      MATRIX.
       CALL CORAN(DATA,NSUB,NDAY,1,0,XBAR,CORR,PCOR,25,15)
       DO107I=1,NDAY
       WRITE(7,26)(CORR(I,J),J=NDAY,I,-1)
26     FORMAT(1H ,15F5.3)
107    CONTINUE
C      CALCULATE AND WRITE VARIANCE FOR EACH LEVEL OF IDAY WITHIN EACH
C      LEVEL OF IDEN.
       CALL CORAN(DATA,NSUB,NDAY,0,0,XBAR,CORR,PCOR,25,15)
       WRITE(6,201)(CORR(M,M),M=1,NDAY)
201    FORMAT(1H0,'VARIANCES:',/,15F7.2)
       L1=1
       L2=15
       DO101ISUB=1,NSUB
       K=0
       DO104JSUB=1,NSUB
       IF(JSUB.EQ.ISUB)GOTO104
       K=K+1
       DO102IDAY=1,NDAY
       DDATA(K,IDAY)=DATA(JSUB,IDAY)
102    CONTINUE
104    CONTINUE
       N=NSUB-1
       CALL CORAN(DDATA,N,NDAY,0,0,XBAR,DCORR,PCOR,25,15)
       DO106I=1,NDAY
       L=(ISUB-1)*NDAY+I
C      CALCULATE JACKKNIFE VARIANCE ESTIMATES.
       Y(L)=NSUB*ALOG(CORR(I,I))-(NSUB-1)*ALOG(DCORR(I,I))
106    CONTINUE
```

```
C     WRITE JACKKNIFE VARIANCE ESTIMATES.
      WRITE(7,27)(Y(I),I=L1,L2)
27    FORMAT(1H ,10F7.2,/1H ,5F7.2)
      L1=L1+15
      L2=L2+15
901   CONTINUE
808   CONTINUE
      STOP
      END
```